

APPLICATION OF DATA MINING-BASED AFFINITY PROPAGATION CLUSTERING ALGORITHM FOR DIAGNOSIS OF MECHANICAL EQUIPMENT TRANSMISSION SYSTEM

Xiaoyi Zheng¹, Peijiang Li¹, Costas Papadimitriou²

¹Quzhou Vocational Technology, Zhejiang 324000, China

²Department of Mechanical Engineering, University of Chessaly, Volos, Greece

Email: mjguo8483@126.com

Abstract – Data mining techniques used to monitor and diagnose the faults of the transmission system of mechanical equipment, thereby promoting the development of big data analysis in the field of intelligent diagnosis. The Affinity Propagation (AP) clustering algorithm is commonly applied in cluster analysis techniques and used as a base to extract the fault information of helicopters. First, the data points are extracted randomly to build a dataset. Then, multiple feature definitions are utilized to extract fault information, which is imported into the AP algorithm for the construction of an early fault clustering effect diagram to judge the early faults of the transmission system and the characteristics of the traditional AP algorithm. Second, the Euclidean distance is weighed negatively; the similarity of the negative Euclidean distance metric is improved and applied to the traditional AP algorithm. The method mentioned above also randomly extracts data points to extract feature information, thereby building a multi-fault diagnosis method for helicopter transmission bearings. Finally, by integrating online and offline data analysis, a fault diagnosis treatment plan for the overall big data is proposed. The results show that, for the early diagnosis of slight changes in data, the application of the AP algorithm can reduce the data dimension and complete data division; however, the separation of data is not ideal, and data overlap occur. After improvement, the differences between the calculation dimension can be presented, while the original characteristics of the data are retained. The effectiveness and resolution of the AP algorithm are improved. The analysis of online and offline data is combined to make the final diagnosis result more accurate and reliable. Reduced number of calculations provide accumulation and support for the fault diagnosis of the transmission system in the future.

Keywords: Data Mining, AP clustering Algorithm, Mechanical Equipment, Transmission System, Fault Diagnosis.

1. Introduction

Cluster analysis technology is one of the important methods of data mining [1, 2]. The clustering analysis is to find the representative points of each class to complete the clustering process. The algorithms of cluster analysis mainly include K-Means, C-Means, and Fuzzy C-Means (FCM) clustering. The Affinity Propagation (AP) clustering algorithm proposed by Frey et al. has been successfully applied to various fields, such as image processing, information retrieval, and medical diagnosis, due to its excellent characteristics, including high operation efficiency, high calculation accuracy, and unlimited similarity matrix. The clustering process is shown in Figure 1. The AP algorithm treats all data objects as a network of nodes and is not affected by the initial cluster center. It recursively transfers real-valued information

between data objects. It calculates the center point of the optimal data; finally, it achieves the ideal clustering effect and extracts the required feature information. However, the AP clustering analysis technique also has shortcomings. For example, Wang et al. proposed that it does not apply to multi-subclass data sets, and the clustering effect is not ideal. For this reason, it is worthwhile to expand and study the AP algorithm to improve its clustering effect.

For mechanical equipment, the transmission system is one of the important components [3]. For example, the working status of the transmission chain in a wind turbine is related to the normal operation. Also, the traction drive system of high-speed trains is the key to ensuring a safe and stable operation. Besides, in the coal mining process, the chain conveyor system of the scraper conveyor directly affects the transport performance and

lifetime. Such examples are countless. Therefore, the establishment of an effective fault diagnosis system to accurately and effectively capture fault characteristics has great scientific significance and application value in terms of the normal operation of mechanical equipment, as well as the reduction of operation and maintenance costs [4]. Helicopters are widely used in the military field. In recent years, they have become more common in daily life. The future development potential of helicopters is huge; however, at the same time, the number of failures and risks accompanying them is also increasing. Among the three key components of a helicopter include the power system, the operating system, and the transmission system, the first two components have corresponding safe redundant designs. In addition, the transmission system requires directional transmission and transmission efficiency of power. With high requirements, part of its safety margin is lost, resulting in its smallest safety redundancy among the three major components of the helicopter. If the transmission system fails, the power transmission to the rotor and tail rotor also fails, which will cause the helicopter to fly abnormally. Thus, the transmission system has a great impact on the safety performance of the helicopter. The performance index of the transmission system in the helicopter is closely related to flight safety. Due to the huge amount of data generated during flight, the aerial fault diagnosis of the helicopter transmission system becomes complicated. Under the background of complex and diverse information, traditional fault diagnosis measures can no longer meet the requirements. At the same time, intelligence and integration will be the future development direction of helicopters. Therefore, the research on fault diagnosis of a helicopter transmission system is vital.

Based on the AP clustering algorithm, this study selects the helicopter transmission system as the research object and establishes helicopter fault diagnosis schemes of different dimensions. By observing the final clustering effect, as well as comparing and analyzing the traditional and improved AP algorithm, the application of the AP clustering algorithm in transmission system fault diagnosis is expanded. At the same time, hoping to provide a reference for the development of cluster analysis technology in the field of intelligent diagnosis in the context of big data.

2. Method

2.1 Fault feature data extraction

For judging whether the working state of the transmission system is normal, collecting data changes is the most direct and effective method. The analysis of the frequency domain signals collected from the helicopter can effectively determine its internal working status, thereby achieving the identification and evaluation of the faults [5]. However, when using big data analysis to deal with failures, it is necessary to judge from multiple angles, which requires redefining the data. In this study, a dimensionless index based on frequency domain analysis [6] is selected. For example, kurtosis combined with skewness can be used to characterize faults; the magnitude and frequency of low-frequency loads can be used to judge faults. The combination of these multiple features can achieve a better discrimination effect. Euclidean distance is also called the Euclidean metric. Negative Euclidean distance is used to calculate the distance between feature sequences, as shown in Equation 1. Defining multiple features similar to the aforementioned as data input points can characterize different aspects of the faults, which is algorithmically reflected in the data dimension.

For the actual fault problem, first, the similar distance between these data points is found and input into the AP algorithm to obtain the feature center of all data points through the algorithm. Then, the most typical features are selected to achieve the purpose of reducing the data dimension, thereby completing the feature clustering. The traditional AP algorithm clustering process is shown in Figure 1.

When extracting feature information, the involvement of massive data points will greatly increase the amount of calculation. Therefore, this study randomly extracts data points to build a dataset and judges the occurrence of the fault from the collected data signals.

$$\partial(d, j)_t = -P\alpha_{dt} - \alpha_{jt} P^2 \quad (1)$$

Where: $\partial(d, j)_t$ represents the sum of the squares of the negative Euclidean distances between the two corresponding feature vectors at time t.

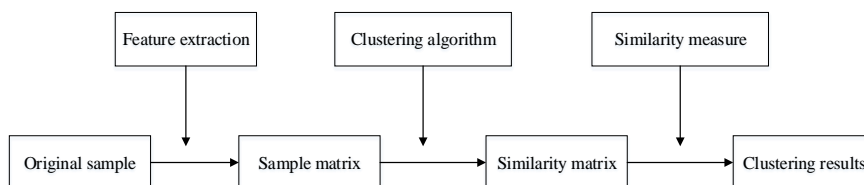


Figure 1: Clustering process of traditional AP algorithm

2.2 Improvements of AP algorithm

Under the effect of the square, the negative Euclidean distance representation increases the difference between the data with large variability, which may lead to unsatisfactory separation results, insufficient data clarity, insufficient accuracy of classification, and poor fault diagnosis. By improving the similarity of the negative Euclidean distance metric used in the traditional AP algorithm, a weighted negative Euclidean distance is used in the AP algorithm to evaluate the impact of certain feature data information on the overall measurement characteristics and achieve the labeling of individual data and dimensions.

Thus, it can better characterize the degree of difference between dimensions, highlight relevant feature data, and make the clustering effect better, thereby showing better results in fault diagnosis, as shown in Equation 2. The clustering flowchart is shown in Figure 2.

$$\partial(d, j)_{i'} = -P \varepsilon_{dt} \alpha_{dt} - \varepsilon_{dt} \alpha_{jt} P^2 \tag{2}$$

Where: ε_{dt} indicates the added weight of each feature data. The influence of the corresponding data on the entire feature set is measured to improve the clustering result of the AP algorithm.

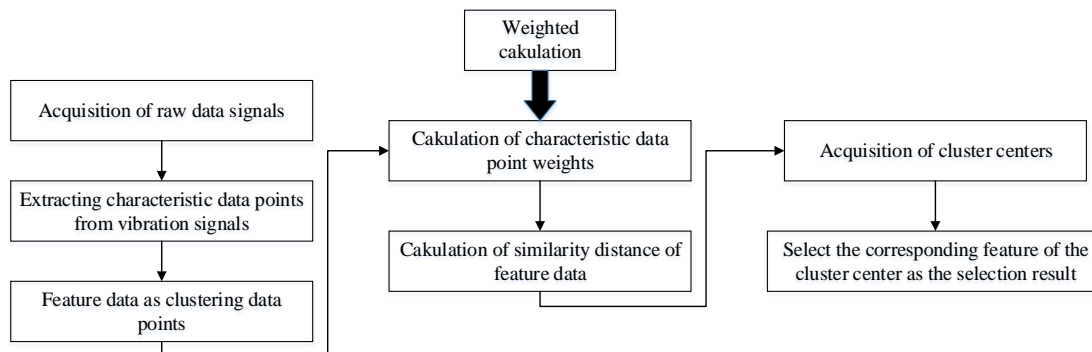


Figure 2: Improved AP algorithm clustering process

2.3 Fault diagnosis scheme

In the fault diagnosis research of the helicopter transmission system, if the corresponding fault can be eliminated in the initial stage of fault occurrence, it will be of considerable significance to improve the safety performance of the machine and avoid the emergence of unexpected conditions. However, when it corresponds to the actual situation, the occurrence and expansion of faults are often the cumulative result of a series of external causes rather than caused by a simple factor. At this time, based on the analysis of a single factor, it is impossible to solve the problem by removing the hidden troubles fundamentally. Research on the direction of multi-fault diagnosis is critically important to find problems more efficiently and to find all existing faults.

Faults caused by multiple factors are difficult to be determined through one or two data analyses. The analysis of massive data can better troubleshoot problems, and the overall analysis will be completer and more effective. With the help of data analysis, accurate classification of multiple mixed faults can be achieved. Therefore, in this study, the AP clustering algorithm based on data mining technology is applied to the fault diagnosis of the helicopter transmission system. From the

perspective of early fault diagnosis, multiple fault diagnosis, and overall fault diagnosis, the application and development in helicopter fault diagnosis of both the traditional AP algorithm and the improved AP algorithm are studied. Among them, the related processing methods for big data fault diagnosis are shown in Figures 3-5.

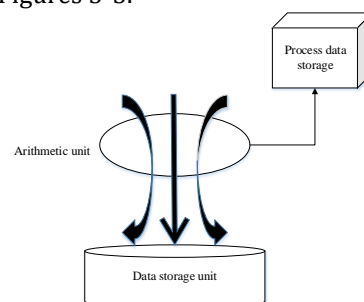


Figure 3: Online data processing flow

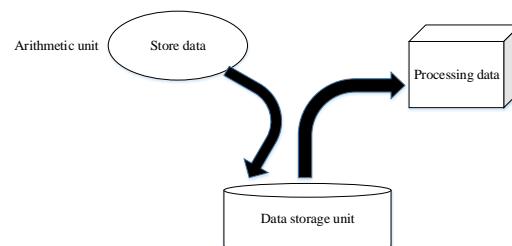


Figure 4: Offline data processing flow

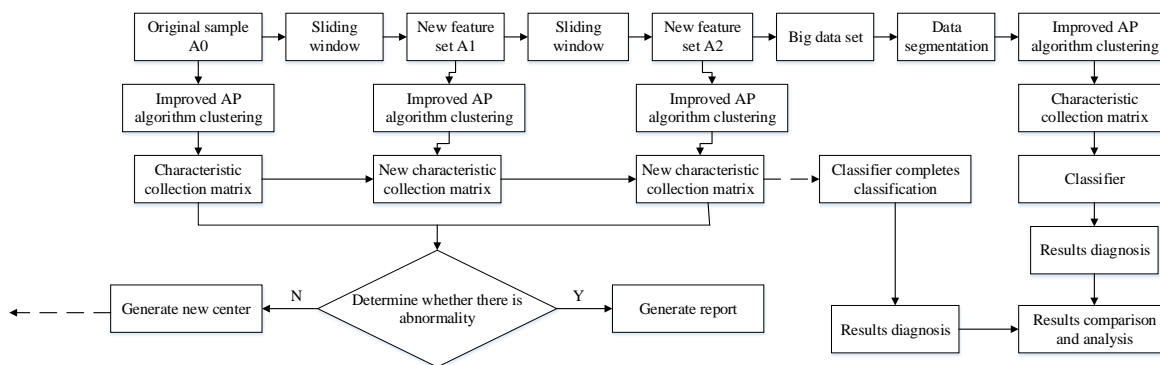


Figure 5: Big data overall fault diagnosis and processing method

3. Results and Discussion

3.1 Early fault diagnosis

The data points are randomly extracted to establish a dataset. The feature data are imported into the AP clustering algorithm for feature extraction, and the traditional AP algorithm is applied to the early diagnosis of the helicopter transmission system to obtain the early fault clustering effect diagram, as shown in Figure 6 below.

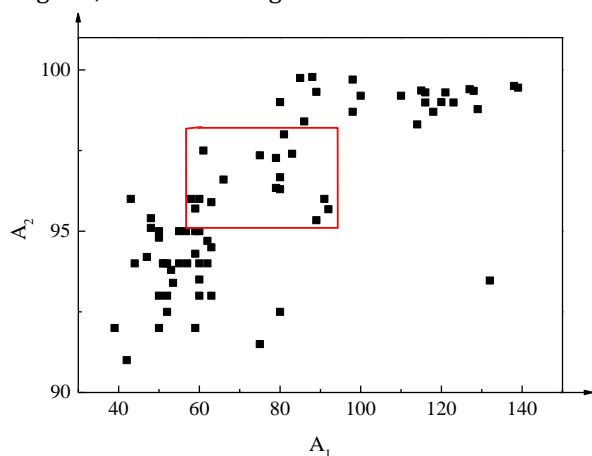


Figure 6: Clustering effect of early fault diagnosis based on traditional AP algorithm

As shown in Figure 6, the effect of the traditional AP algorithm in the early fault diagnosis of the helicopter transmission system is quite apparent. It can accurately extract the specific information, effectively reduce the data dimension during the operation, and play a role in fault differentiation. It can solve problems that are difficult to find in early failures and can play an active role in early failure diagnosis and simple failure diagnosis. However, it can be found from the clustering effect that there is an overlapping area of data (the area shown in the red box in the figure), and the separation result is not very satisfactory. The results show that the traditional AP algorithm in helicopter transmission system fault diagnosis application has the problems of insufficient clarity between data clusters and poor separation effect. If the amount of fault data is large

and the data changes are significant, it may cause misjudgment of faults, which will affect the normal working status of the entire transmission system. Thus, the traditional AP algorithm has defects and is not suitable for application in complex situations.

In the helicopter transmission system, under normal circumstances, the early failure of ordinary parts is not apparent, and the corresponding data changes are small. However, with the extension of the working time limit, the probability of sudden changes caused by it will increase, and the future risk factor also increases. For this reason, the early diagnosis of faults is very critical.

3.2 Multi-fault diagnosis based on improved AP algorithm

Based on the analysis methods and results of early fault diagnosis, the weighted negative Euclidean distance is used to improve the AP algorithm. The method of randomly selecting data points is also utilized to establish a feature dataset and complete the clustering. Besides, the weighted dimensions are used to calculate the differences between dimensions to solve the problem of helicopter transmission bearing failure. Combined with the K-Means algorithm, the feature data is imported; the K-Means algorithm is used to complete the clustering, and the clustering results of the improved AP algorithm are compared.

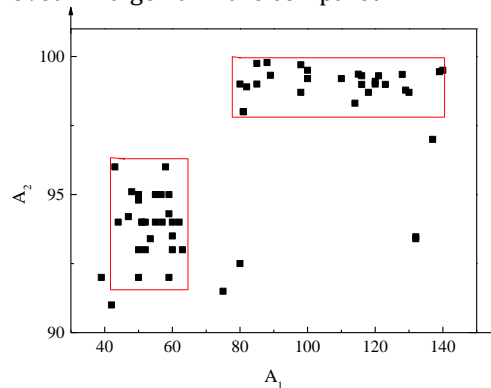


Figure 7: Clustering effect of multi-fault diagnosis based on improved AP algorithm

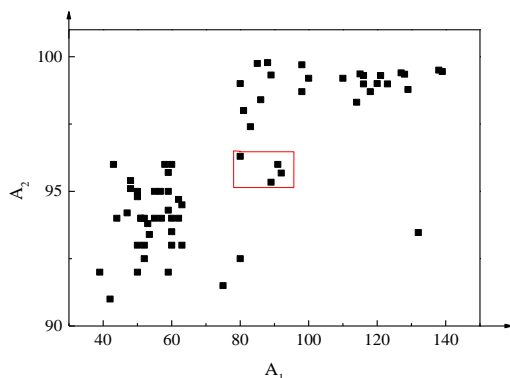


Figure 8: Clustering effect of multi-fault diagnosis based on K-Means algorithm

The multi-fault diagnosis results for bearings show that the original characteristics of important dimensional data are still retained, and the characteristics are prominent. The analyzable data becomes larger and the clustering effect is better. As shown in the red box in the figure, the data classification is visible. Compared to the traditional AP clustering algorithm, the improved AP clustering algorithm has a more significant effect on fault diagnosis, which can clearly distinguish fault feature information. Even for fault data with a high degree of similarity to normal data, it still has a good distinction. As the difference between distance and space becomes larger, the discrimination is more accurate. Compared to the traditional AP algorithm, it has a better effect on the problem that early faults are difficult to find. For the effect of the K-Means algorithm clustering, the red area indicates that there is data overlap. In contrast, it can be found that the improved AP algorithm can accurately distinguish, thereby the calculation efficiency is higher, and the separation effect is better.

Rolling bearings are one of the most easily damaged parts of rotating machinery. Whether their working conditions are normal and whether they can operate normally has a direct impact on the performance, efficiency, and life of the entire mechanical equipment. Therefore, the fault diagnosis of the bearing is necessary [7, 8]. If a bearing failure occurs in the transmission system, it will cause wear of the inner ring, disconnection of the cage, and severe vibration of the entire transmission shaft. It may even cause serious consequences such as transmission failure because the transmission shaft is not controlled. Many experts and scholars have researched the normal operation of rolling bearings and the means to maintain a normal state. For example, Kim et al. [9] proposed a method for assessing bearing life by using the Support Vector Machine (SVM) classifiers. Also, scholars from Xi'an Jiaotong University have researched feature extraction and equipment performance during weak faults. The collection of bearing data is generally generated from the entire component; therefore, it is

difficult to determine the specific location. Also, because the amount of data involved is huge and complex, if it continues to select a simple fault diagnosis method, the fault cannot be eliminated from the root cause. Thus, this study uses multi-fault diagnosis methods. As can be inferred from the above fault diagnosis analysis of helicopter transmission bearings, the improved AP algorithm can well complete the clustering of the data set for the mechanical equipment with precise and complicated structure. At the same time, it can clearly classify the fault category for the diagnosis of the different parts of the fault in the same equipment, which is more accurate and more reliable. It shows that the application of the improved AP method is effective for high-precision and high-demand equipment fault diagnosis in helicopter transmission systems. This can solve the difficult problem of precise equipment fault diagnosis in the helicopter transmission system, and this application can also be extended to the fault diagnosis of other mechanical equipment transmission systems [10-14].

3.3 Overall fault diagnosis based on the improved AP algorithm

By integrating online fault diagnosis and offline fault diagnosis, as well as integrating the timeliness of online data processing and the accuracy of offline data processing, this study applies the improved AP clustering algorithm to the overall big data fault diagnosis. Based on the improved AP clustering algorithm, for online data, a sliding window method is used to select a data stream at a specific period as the analysis sample, and the improved AP algorithm is used to cluster the samples. Changes need to be reported, and those within the rules will continue to be updated to complete the clustering. As new samples re-appear as the window slides, it continues to cluster and obtain the cluster center and compares and determines whether there are differences. In this way, the cycle is scrolled until the end of the data stream. Thus, the required fault feature information is extracted. The online data processing flow is shown in Figure 3. For the fault diagnosis and analysis of offline data, to ensure the calculation speed and accuracy, the huge amount of data is segmented, data mining is performed from each small segment, and the improved AP algorithm is used for clustering and processing the offline data. The process is shown in Figure 4. The online and offline data diagnosis is combined to establish an overall big data fault diagnosis and processing method, as shown in Figure 5.

Taking the transmission gear as the research object, an overall fault diagnosis scheme for the outer ring diameter is established, and the fault data is analyzed. The clustering effect is shown in Figure

9, where different red dots in the figure represent fault data corresponding to different gear diameters. As shown in the clustering effect diagram, after clustering, the fault separation result is still relatively obvious. The overall fault of the transmission system is complex, and the existence of overlapping parts is consistent with the actual situation, which indicates that gears with different diameters may have similar fault performance. Therefore, the fault diagnosis scheme is effective for the overall complex fault diagnosis, which can identify complex faults, meet the needs of actual engineering, and solve the problem of unclear fault diagnosis.

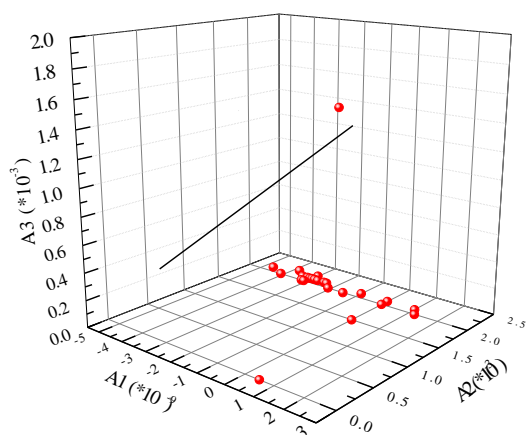


Figure 9: Clustering effect of overall fault diagnosis based on the improved AP algorithm

The online data diagnosis and processing can analyze the data during the data collection process. It can complete the real-time monitoring of the relevant working parameters during the use of the transmission system, as well as timely diagnosing and reporting the possible failure information. Also, the offline data diagnosis and processing are for the database. The collected overall data is analyzed, and useful information is mined for results extraction. Combining online fault data processing, offline fault data processing, and data analysis, a comprehensive significant data fault processing solution is obtained. For the online fault data processing results, the final diagnosis result is obtained through the secondary classification of the classifier, which is combined with the offline fault diagnosis. In addition to comparing and analyzing the processing results, as well as the real-time and accurate detection of failures of the transmission system of mechanical equipment, big data can also be used to mine potential failure signs and predict the service life of the transmission system.

In the actual use process, the occurrence of mechanical equipment failure generally corresponds to a massive amount of data instead of a small number of data analysis that cannot be reflected and characterized. To make a better prediction and

control system in advance, it is necessary to analyze the enormous amount of data to find the problem faster and eliminate the source of the fault, thereby ensuring that the mechanical equipment transmission system is in a healthy and stable working state. Therefore, it is significant to integrate online data processing and offline data processing to establish a method for the overall fault diagnosis and processing of big data.

4. Conclusions

In this study, the fault diagnosis of the helicopter transmission system is used as the starting point. The application of the AP clustering algorithm in the fault diagnosis of the mechanical equipment transmission system is studied. Based on the traditional AP clustering algorithm, early fault diagnosis is established. The AP clustering algorithm is improved based on weighting to establish a multi-fault diagnosis for rolling bearings, with an online and offline integrated big data fault diagnosis scheme.

The analysis results show that compared to the traditional AP clustering algorithm, the improved AP clustering algorithm has distinct advantages in terms of data calculation rate, separation effect, and accurate identification. Due to the ability to accurately extract characteristic target information, the traditional AP clustering algorithm can be applied in early fault diagnosis. However, the clustering effect is still worse compared to the improved AP clustering algorithm. In the multi-fault diagnosis of transmission bearings, the improved AP clustering algorithm can divide the fault categories, which is accurate and reliable. The timeliness of online data analysis and the accuracy of offline analysis are integrated. The improved AP clustering algorithm plays a significant role in the overall big data fault diagnosis. The prediction of the service life of the transmission system provides a particular reference value.

References

- [1] Luo R, Wang J, Zhong H, et al. SU-F-R-33: Can CT and CBCT Be Used Simultaneously for Radiomics Analysis. *Medical Physics*, 2016, 43(6), pp. 3380-3380.
- [2] Siebert U. Population Differentiation and Climatic Adaptation for Growth Potential of White Spruce (*Picea glauca*) in Alberta. *Silvae Genetica*, 2017, 59(1-6), pp. 158-169.
- [3] Kosuge A, Hashiba J, Kawajiri T, et al. An Inductively Powered Wireless Solid-State Drive System With Merged Error Correction of High-Speed Wireless Data Links and NAND Flash

- Memories. IEEE Journal of Solid-State Circuits, 2016, 51(4), pp. 1-10.
- [4] S.K. Gugulothu, N. Prabhu Kishore, V. Phani Babu, Girish Sapre (2019). Computational Investigation of Mhd Free Convection on A Flat Vertical Plate Embedded with Micro Polar Fluid. Acta Mechanica Malaysia, 2(1): 16-22.
- [5] P. Umamaheswarrao, D. Ranga Raju, Kns Suman And B. Ravi Sankar (2019). Topsis Based Optimization Of Process Parameters While Hard Turning Of Aisi 52100 Steel. Acta Mechanica Malaysia, 2(2): 28-31, DOI: 10.26480/amm.02.2019.28.31
- [6] Hesham Enshasy, Qasem Abu Al-Haija, Hasan Al-Amri, Mohamed Al-Nashri, Sultan Al-Muhaisen (2019). A Schematic Design of HHO Cell as Green Energy Storage. Acta Electronica Malaysia, 3(2): 09-15.
- [7] Hemad Heidari Jobaneh (2019). An Ultra-Low-Power and Ultra-Low -Voltage 5 GHz Low Noise Amplifier Design with Precise Calculation. Acta Electronica Malaysia, 3(2): 23-30.
- [8] Bawar Mohammed Faraj and Faraedoon Waly Ahmed (2019). On The Matlab Technique By Using Laplace Transform For Solving Second Order Ode With Initial Conditions Exactly. Matrix Science Mathematic, 3(2): 08-10.
- [9] Yang D, Liu Y, Li X, et al. Gearbox fault diagnosis based on bacterial foraging algorithm optimization decisions. Journal of Central South University, 2015, 46(4), pp. 1224-1230.
- [10] Zhou M, Wang Z H, Shen Y. Fault detection and isolation method based on H_2/H_∞ unknown input observer design in finite frequency domain. Asian Journal of Control, 2017, 19(5), pp. 1777-1790.
- [11] Martins G, Cordioli J A. A computationally efficient method for the frequency-domain analysis of visco-thermal acoustic propagation in arbitrary geometries. Journal of the Acoustical Society of America, 2015, 138(3), pp. 1828-1828.
- [12] Glowacz A, Glowacz W, Glowacz Z, et al. Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals. Measurement, 2017, 113, pp. 1-9.
- [13] Abdelkader R, Kaddour A, Derouiche Z. Enhancement of rolling bearing fault diagnosis based on improvement of empirical mode decomposition denoising method. International Journal of Advanced Manufacturing Technology, 2018, 97(13), pp. 1-19.
- [14] Kim H E, Tan A C, Mathew J, et al. Bearing Fault prognosis based on health state probability estimation. Expert Systems with Applications, 2015, 39 (5), pp. 5200-5213.